

Previsão de risco de alagamento com uso de inteligência artificial

Sabrina Sambati, Renata Genova Martins, Ricardo Batista Vilela, Alex Sandro Aguiar
Pessoa, Danielle de Almeida Bressiani

Climatempo - Agência Brasileira de Meteorologia Ltda., São José dos Campos, SP
sabrina.sambati@climatempo.com.br; danielle.bressiani@climatempo.com.br

Abstract: A transição para cidades inteligentes e resilientes é uma estratégia para auxiliar na prevenção e mitigação dos danos relacionados a eventos hidro-meteorológicos severos, que frequentemente acometem a Região Metropolitana de São Paulo (RMSP). Foi aplicado *k-Nearest Neighbor*, um algoritmo de aprendizado de máquina, na previsão de ocorrência de alagamentos e inundações. Foi escolhida a RMSP como área de estudo por representar a área com maior aglomeração urbana do Brasil e receber influência drástica de eventos meteorológicos severos e seus impactos.

1. Introdução

A situação de riscos ambientais e seus impactos no Brasil [1], Estado de São Paulo [2] e Grande São Paulo [3] é alarmante. Áreas urbanas com elevada concentração populacional possuem alta vulnerabilidade sociais e ambientais. A Região Metropolitana de São Paulo (RMSP), uma das maiores aglomerações urbanas do mundo, sofre com eventos meteorológicos severos pois provocam frequentes transtornos a população decorrente das inundações e deslizamentos.

O conceito cidades inteligentes originou-se para alcançar os objetivos estabelecidos pelo Protocolo de Kyoto [4] e sua adoção é uma estratégia para a redução da vulnerabilidade urbana, atuando nas necessidades da comunidade com o emprego de novas tecnologias para contribuir com o processo de suporte à decisão, viabilizando a participação ativa do cidadãos.

No âmbito do projeto de pesquisa 2016/10229-3, financiado pela FAPESP e FINEP, foi desenvolvido um sistema para auxiliar cidadãos e empresas a lidar com riscos ambientais na RMSP, sendo que a interface principal desse sistema com o público será o aplicativo Pedágua [5]. Neste produto foi desenvolvido um sistema que constrói campos de risco de alagamentos e inundações baseados em dados de precipitação em tempo real; de previsão de curtíssimo prazo (*nowcasting*); e mapeamento de vulnerabilidade à alagamentos e inundações combinados através de técnicas de inteligência artificial (IA).

Técnicas de IA vem sendo muito exploradas pelas mais diversas aplicações, principalmente devido ao crescimento exponencial das bases de dados e pelo conhecimento nela existente. IA passou a auxiliar os analistas, pesquisadores e usuários a tomarem decisões com mais clareza, rapidez e assertividade. O aprendizado de máquina, uma área de IA, permite aos computadores adquirirem conhecimento com a experiência, exemplo e analogia [8]. Dentre os diversos algoritmos de aprendizado de máquina existentes, o kNN – (*k-Nearest Neighbor*) é um dos mais antigos e simples para classificação de padrões, foi proposto por [9] e é considerado um método de aprendizado não paramétrico, já que os dados não precisam estar normalmente distribuídos e nem necessita de aprendizado supervisionado.

Este estudo mostra a aplicação do kNN na identificação de alagamento e inundação na Região Metropolitana de São Paulo através do uso de diferentes combinações de dados ambientais.

2. Metodologia

Para o desenvolvimento desta pesquisa a metodologia foi dividida em algumas etapas conforme a Figura 1.

2.1. Previsão com kNN

O kNN – (*k-Nearest Neighbor*) é adequado para tarefas de classificação, sendo que o aprendizado tem como base a analogia. Para determinar a classe de um novo elemento, é utilizado uma base de referência cujos dados já estejam rotulados (classificados) onde o kNN calcula a distância entre essa nova ocorrência e todos os dados pertencentes a base

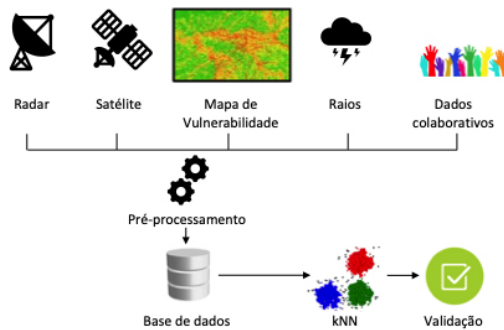


Figura 1. Fluxograma das etapas da pesquisa

de referência usando uma métrica onde é criado um vetor com as respectivas distâncias. Os k primeiros elementos desse vetor são selecionados (mais similares) e o rótulo mais frequente entre os elementos é atribuído a nova ocorrência.

Os principais dados utilizados para montar a base de referência foram: precipitação estimada por dois radares meteorológicos banda X do projeto Chuvaonline (da USP e Climatempo, com resolução espacial de 100 m e temporal de 5 minutos); precipitação estimada a partir dos dados do satélite GOES-16 (com resolução espacial de 2 km e temporal de 15 minutos); mapeamento de vulnerabilidade física da RMSP, os dados colaborativos foram usados como rótulos: ocorrência ou não de alagamento.

Para determinação do mapa de chuva, foi utilizado o acumulado de precipitação das últimas 2 horas sob as áreas das bacias hidrográficas separadamente, assim como as intensidades das precipitações. Com base nessas duas informações foi calculado um mapa de chuva integrando os dados da chuva estimada através dos radares e do satélite GOES-16. Já para a confecção do mapa de vulnerabilidade foram utilizados 5 parâmetros físicos (como impermeabilização do solo, proximidade aos canais, declividade, área de drenagem, ocorrência passada de alagamentos) da região que foram determinados e priorizados por especialistas de acordo com a metodologia AHP.

Os dados colaborativos utilizados são da base de dados do Centro de Gerenciamento de Emergências de São Paulo (CGE) e oriundos de dados colaborativos obtidos em mídias e plataformas sociais. Os dados do CGE são coletados diretamente pelo site do órgão. Já o processo de obtenção dos dados colaborativos é realizado em três fases: i) um coletor de dados ambientais que reúne as frases obtidas pelos relatos das diversas fontes (*Twitter*, *Flickr*, *Instagram* e outros). ii) estas informações passam por um algoritmo que filtra as mensagens georreferenciadas através de processamento de linguagem natural e agrupa (através do método de estimativas de densidade de *kernel*) os relatos próximos no espaço e tempo, sendo que para ocorrer o agrupamento é necessário considerar um mínimo de três relatos, a uma distância máxima entre eles de 200 m , em intervalo de tempo de 30 minutos; iii) uma análise de falsas ocorrências é realizada por meio da verificação, em tempo real, da chuva estimada por radares ou satélite na bacia hidrográfica da ocorrência. Detalhes adicionais da metodologia podem ser obtidas no trabalho de [10] e [11].

Depois do desenvolvimento de toda a base de dados, antes da entrada para o kNN, os mesmos passaram por um pré-processamento [12] consistindo das seguintes fases: i) limpeza dos dados: uniformização dos dados, eliminando as inconsistências; ii) integração dos dados: agrupando os dados de diferentes repositórios; formando um único; iii) transformação dos dados: padronização dos dados em categóricos, numéricos; iv) redução dos dados: restringindo o volume de dados para um intervalo adequado.

A base de dados completa, resultante do pré-processamento, composta por 18 atributos que correspondem as variáveis meteorológicas, mapa de vulnerabilidade e dados colaborativos, listados na Tabela 1 foi normalizada através do método padrão (max e min).

Para este trabalho, foi realizado um estudo de caso cujo recorte temporal adotado para testes foi o mês de janeiro de 2018 e para validação o dia 17 de outubro de 2018 (a escolha desta data é justificada devido ao relato de diversos alagamentos na RMSP [13]). Com intuito de analisar a influência de diferentes parâmetros e atributos na previsão de alagamento pelo kNN, alguns cenários foram criados variando o valor de k (único parâmetro livre do kNN) e o conjunto de atributos (sub-conjuntos da base de referência) selecionados pelo WEKA – *Waikato Environment for Knowledge Analysis*. Para selecionar os melhores cenários a avaliação foi feita visualmente e qualitativamente (número de pontos de alagamentos registrados).

Tabela 1. Atributos usados no kNN.

Latitude	Longitude
Vulnerabilidade	Alagamento
Precipitação na bacia – 1h (Radar)	Precipitação na bacia – 2h (Radar)
Precipitação na bacia – 3h (Radar)	Precipitação na bacia – 24h (Radar)
Precipitação na bacia – 1h (Hidroestimador)	Precipitação na bacia – 2h (Hidroestimador)
Precipitação na bacia – 3h (Hidroestimador)	Precipitação na bacia – 24h (Hidroestimador)
Raio – 60 min (<i>Starnet</i>)	Raio – 60 min (<i>Earth</i>)
Precipitação – 1h (Radar)	Precipitação – 1h (Hidroestimador)
Taxa precipitação (Radar)	Taxa precipitação (Hidroestimador)

3. Resultados

Após uma análise dos testes realizados foi possível observar que o desempenho do kNN está fortemente ligado ao conjunto de atributos usado. A Figura 2 corresponde aos resultados obtidos com 3 diferentes cenários (conjuntos de atributos diferentes) e com o valor de $k = 1$. Na fig.2-a o conjunto utilizado foi: Raio – 60 min (*Earth*), Taxa precipitação (Hidroestimador), Precipitação – 1h (Hidroestimador), Precipitação na bacia – 3h (Hidroestimador), Precipitação na bacia – 24h (Hidroestimador); já na fig.2-b foi: Precipitação – 1h (Radar), Precipitação – 1h (Hidroestimador), Precipitação na bacia – 1h/2h/3h/24h (Hidroestimador), Precipitação na bacia – 1h/2h/3h/24h (Radar) e na fig.2-c o conjunto de atributos utilizados foi: Raio – 60 min (*Earth*), Precipitação – 1h (Radar) e Precipitação na bacia – 1h (Hidroestimador).

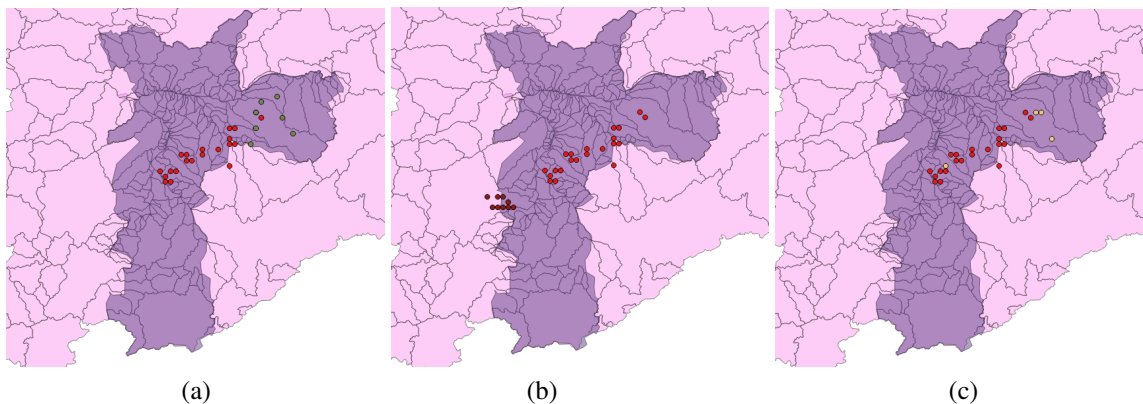


Figura 2. Resultado do kNN com diferentes cenários

Observa-se que os pontos vermelhos representam os dados de alagamento observados (a partir dos dados colaborativos - lembrando que há uma limitação na disponibilidade destes dados) e os pontos coloridos equivalem aos resultados do kNN, cada cor representa um cenário distinto. sendo o fig.2-c correspondente ao melhor resultado. Apesar dos testes serem preliminares e necessitarem de ajustes, é interessante observar que eles já seguem um padrão similar ao observado.

Um outro ponto a ser observado é que variáveis importantes para identificação de alagamento não foram selecionadas no conjunto de atributos montados pelo WEKA (ex. vulnerabilidade física do terreno). Demonstrando portanto, a necessidade de utilização de outros métodos de composição de conjuntos e a necessidade de uma base de referência de teste significativa.

4. Conclusão

As técnicas de inteligência artificial tem mostrado eficiência em várias aplicações meteorológicas e ambientais que exigem computação intensiva e precisão, pois proporcionam uma diminuição considerável do esforço computacional garantindo resultados eficientes e precisos, o que é muito adequado em casos que exigem respostas imediatas, como é o caso desse projeto. Os resultados mostram que a abordagem proposta usando o kNN, um método com implementação

simples, apresenta resultados promissores, porém percebe-se que a fase de ajustes é a que exige mais dedicação, o que é esperado quando se utilizam técnicas de IA, já que os resultados preliminares mostraram que o desempenho do kNN está diretamente ligado ao conjunto de atributos utilizado. O método será mais amplamente testado e validado. Sabe-se que o kNN é um método computacionalmente caro quando aplicado em base de dados grandes, mas isso não foi levado em consideração nesse trabalho, pois o foco deste estudo foi validar o comportamento da metodologia adotada na previsão de alagamento e inundação.

5. Referências

- [1] Atlas Brasileiro de Desastres Naturais: 1991 a 2012 / Centro Universitário de Estudos e Pesquisas sobre Desastres, Florianópolis: CEPED UFSC, (2013).
- [2] Brollo M J, Tominaga L K , “Desastres naturais e riscos geológicos no Estado de São Paulo: cenário de referência”, Boletim nº 1, Grupo de Articulação de Ações, , 2012.
- [3] Silva-Dias M A F DA et al., “Changes in extreme daily rainfall for São Paulo, Brazil”, *Climate Change* , 705–722, (2013).
- [4] Annalisa Cocchia, “Smart and digital city: A systematic literature review”, in *Smart city*, 13–43, Springer, (2014).
- [5] Vilela, R. B.; Martins, R. G.; Bressiani, D. A.; Palma, Gilca, “Sistema para previsão de riscos de alagamentos e inundações: um estudo de caso para a Região Metropolitana de São Paulo (RMSP)” in *XVII Encontro dos alunos de pós-graduação em meteorologia do CPTEC/INPE*, Cachoeira Paulista/SP (2018).
- [6] McCulloch, Warren S and Pitts, Walter, “A logical calculus of the ideas immanent in nervous activity”, *The bulletin of mathematical biophysics*, **5**, 115–133 (1943).
- [7] McCarthy, John and Minsky, Marvin L and Rochester, Nathaniel and Shannon, Claude E, “A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955”, *AI magazine* (2006)
- [8] Negnevitsky, Michael “Artificial intelligence: a guide to intelligent systems” (2005).
- [9] Cover, Thomas and Hart, Peter, “Nearest neighbor pattern classification”, *IEEE transactions on information theory*, **13**, 21–27 (1967)
- [10] Flavio Horita; Ricardo Vilela; Renata Martins; Danielle Bressiani; Gilca Palma; João Porto de Albuquerque, “Determining flooded areas using crowd sensing data and weather radar precipitation: a case study in Brazil,” in *15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018)*, Rochester, USA , pp. 1–10.
- [11] Horita, F. E. A.; Vilela, R.; Martins, R. G.; Bressiani, D. A.; Albuquerque, J. P., “Uso de dados colaborativos na determinação de alagamentos.” in *XVII Encontro dos alunos de pós-graduação em meteorologia do CPTEC/INPE* Cachoeira Paulista/SP (2018).
- [12] Han, Jiawei and Pei, Jian and Kamber, Micheline, *Data mining: concepts and techniques*, Elsevier (2011).
- [13] Centro de Gerenciamento de Emergências Climáticas [acessado em Outubro de 2018], <https://www.cgesp.org/v3/alagamentos>.